2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

A Hybrid Transfer Learning Approach to Migratable Disaster Assessment in Social Media Sensing

Yang Zhang, Ruohan Zong, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
{yzhang42, rzong, dwang5}@nd.edu

***Abstract*—Social media sensing has emerged as a powerful sensing paradigm to collect the observations of the physical world by exploring the "wisdom of crowd". In this paper, we focus on a *migratable disaster damage assessment* problem in social media sensing applications. Our goal is to accurately identify the damage severity of affected areas in an *unfolding* disaster event using *unlabeled* social media data feeds (e.g., image posts on social media). Two fundamental challenges exist in solving our problem: i) different disaster events often have distinct characteristics (e.g., damage types, affected areas) that cannot be easily migrated; ii) it is non-trivial to modify a damage assessment model from a previous event to adapt to a new event without using the labeled data from the new event. To address the above challenges, we develop *SocialTrans*, a hybrid deep transfer learning framework, to enable effective model migration for accurate damage assessment *without using any training data* from the studied disaster event. The evaluation results on four real-world disaster events show that SocialTrans consistently outperforms the state-of-the-art baselines in accurately assessing the damage level of disasters.**

I. Introduction

In this paper, we propose a deep transfer learning approach to address the migratable disaster damage assessment problem in social media sensing applications. Social media sensing has emerged as a powerful sensing paradigm to collect the observations of the physical world by exploring the "wisdom of crowd" [1], [2]. Examples of such applications include disease outbreak detection using Twitter feeds [3], intelligent traffic monitoring using Instagram posts [4], and human mobility sensing using Foursquare check-ins [5]. In this paper, we focus on an emerging social media sensing application - *disaster damage assessment (DDA)* [6]. The goal of DDA applications is to automatically assess the damage severity of an affected area from the imagery data posted on social media in the aftermath of a disaster event using AI (e.g., deep learning) techniques [7]. The assessment information can then be used by the emergency response agencies (e.g., FEMA, fire departments) for a timely and effective disaster response.

A good amount of efforts have been made to address the disaster damage assessment problem in data mining, network sensing, and machine learning communities [8]–[12]. Examples include deep convolutional neural network approaches [8], crowd-AI hybrid solutions [11], and deep domain adaptation approaches [12]. Many of these solutions would require a high-quality training dataset (e.g., labeled imagery posts of damage severity) from the disaster event of interest to build an accurate damage assessment model [8]. However, such a high-quality training dataset is often not available for an *unfolding* disaster because of the "cold start" problem [13] and the lack of real-time annotations due to the cost and resource constraints [14]. A recent effort develops a damage detection model that does not require the training data from the studied disaster event. However, such an approach only determines whether an image is damage related or not but fails to assess the *damage severity* of the image [12]. Therefore, it remains as a fundamental challenge to provide an accurate and timely damage assessment of an unfolding disaster using *unlabeled* social media sensing data.

To address the above challenge, this paper focuses on a *migratable disaster damage assessment problem*. The goal is to accurately identify the damage severity of affected areas for a *target disaster event* that is unfolding and does not have any training data by "migrating" a damage assessment model learned from a *source disaster event* that happened earlier where the training data is available. For example, consider two disaster events: Typhoon Ruby in 2015 and Hurricane Matthew in 2016. The social media imagery posts were collected from both events for damage assessment and we only had the training data of Typhoon Ruby at the time when Hurricane Mathew was unfolding. Our objective here is to accurately identify damage severity of affected areas in Hurricane Matthew (target disaster event) by leveraging the damage assessment model obtained from Typhoon Ruby (source disaster event). The above migratable disaster damage assessment problem is non-trivial to solve due to several key technical challenges we elaborated below.

Discrepancy on Event-specific Damage Characteristics. A straightforward solution to solve our problem is to assess the damage of the target event by directly using the model learned from a source event [9]. However, this solution ignores an important fact that the source and target events may have very different damage characteristics. For example, as shown in Figure 1, we observe clear visual differences between different disaster events (e.g., collapsed concrete in (A) vs. broken wooden beam in (B), and flood damage in (C) vs. wind damage in (D)). A recent study also shows a significant accuracy

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

drop when the disaster damage assessment model is directly used across events [8]. The domain adaption solutions can potentially be applied to address the discrepancy challenge by modifying the social media images from the source event to match the damage characteristics of the target event [15], [16]. However, those approaches are often designed for specific computer vision tasks (e.g., identifying numbers from handwriting) and can not handle the complex damage characteristics and excessive fine-grained details of social media images in the DDA application. Therefore, the migratable disaster damage assessment model has to carefully accommodate the discrepancy on event-specific damage characteristics to offer the desirable accuracy for the disaster damage assessment task.

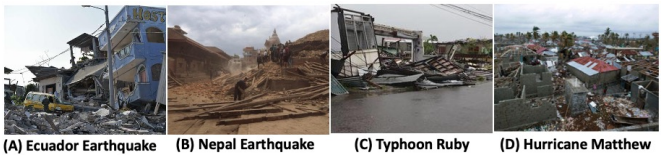


Figure 1. An Example of Event-specific Damage Characteristics

Unsupervised Model Migration between Source and Target Events. An effective way to overcome the discrepancy between the source and target disaster events is to judiciously "migrate" (e.g., modify) the damage assessment model learned from the source event to the target event using transfer learning techniques [11]. In particular, the current solutions identify a critical set of damage features (e.g., structural damage, landscape change) shared between the source and target events and develop migration models that leverage the identified features to estimate the damage severity of images from the target event. However, a major limitation of these solutions is that they require a good amount of training data from the target event to ensure the migration model is sensitive to the damage features from the target event. Unfortunately, such a training dataset is unavailable in the migratable damage assessment problem we focus on in this paper. Therefore, it remains a difficult task to design an effective *unsupervised* migration model without using any training data from the target event.

To address the above challenges, we develop *SocialTrans*, a hybrid deep transfer learning framework to solve the migratable disaster damage assessment problem in social media sensing applications. In particular, we create an adversarial co-training network architecture to enable an effective model migration from the source to the target event by accommodating the discrepancy on event-specific damage characteristics. Furthermore, we design a set of reconstructive and discriminative neural networks to identify the key damage features shared by both source and target events for accurate damage severity level identification. To our knowledge, SocialTrans is the first hybrid deep transfer learning approach that integrates the deep image reconstruction with adversarial training in an end-to-end neural work architecture to solve the migratable disaster damage assessment problem. The results show that SocialTrans consistently outperforms the state-of-the-art baselines by achieving the highest accuracy in correctly identifying the damage severity of affected areas for the target event under different types of disasters.

II. Related Work

A. Social Media Sensing

Motivated by the ubiquitous internet connectivity and information dissemination media (e.g., Twitter, Instagram), social media sensing has emerged as a new approach to obtain a rich set of observations of the physical world at an unprecedented scale [1], [17]. Examples of social media sensing applications including detecting infectious disease outbreak using Twitter feeds [3], predicting real-time traffic risks in cities using Instagram posts [4], and monitoring human mobility in a metro area using online social media check-ins [5]. Several key challenges exist in current social media sensing applications. Examples include data reliability [18], device heterogeneity [19], task allocation [20], incentive design [21], and privacy protection [22]. However, the unsupervised migratable disaster damage assessment remains to be a challenging problem in social media sensing applications. In this paper, we developed the SocialTrans scheme to address this problem by designing a novel hybrid deep transfer learning framework to ensure the desirable migratable damage assessment performance.

B. Disaster Damage Assessment

Previous efforts have been made to address the disaster damage assessment problem in social network analysis, data mining, and deep learning [8]–[12]. For example, Nguyen *et al.* proposed an artificial intelligence driven disaster response system that utilizes social media imagery data for damage assessment by fine-tuning the convolutional neural networks [8]. Li *et al.* utilized the pre-trained deep neural networks to localize and quantify the damages reported on social media images for disaster response [9]. Alam *et al.* developed an end-to-end social media image processing model for damage severity assessment during natural disasters using deep neural networks [10]. However, those approaches cannot be applied to solve our migratable disaster damage assessment problem because they often depend on a high-quality training dataset from the target disaster event, which is not available in our problem setting. Recently, a piece of work that goes closet to ours developed a deep domain adaptation framework that can effectively detect damage related images with minimal training data from the target event [12]. However, the neural network architecture design of their framework is primarily designed for the binary detection task and cannot be easily extended to handle the complex and fine-grained differences in damage characteristics between different damage levels in our DDA application. In contrast, we design an adversarial co-training deep architecture in SocialTrans that integrates the deep image reconstruction, adversarial training, and damage severity classification into a holistic framework to address the above challenge.

C. Deep Transfer Learning

Our work is also related to the deep transfer learning and domain adaptation techniques, which have been used in

many domains such as computer version, natural language processing, image classification, and information retrieval [15], [16], [23]–[27]. For example, Ghifary *et al.* developed a deep transfer learning approach that jointly leverages the deep reconstruction and classification networks for effective migratable image classification tasks [15]. Sankaranarayanan *et al.* designed a generative adversarial network approach to learn the symbiotic relation between the source and target domains for robust image generation [16]. Min *et al.* proposed a context-aware question answering framework that utilizes fine-grained supervision data to improve the answer accuracy through domain adaptation model design [28]. Zhou *et al.* developed a deep domain adaptation framework for sentiment classification of cross-lingual text documents via a weakly shared neural network model [26]. However, those approaches cannot be directly applied to solve our problem as they are often designed for specific tasks (e.g., classifying different types of electronic devices) and cannot effectively address the complex damage characteristics and excessive details of social media images in our DDA application. To the best of our knowledge, SocialTrans is the first hybrid deep transfer learning approach to solve the migratable disaster damage assessment problem in social media sensing applications by integrating the deep image reconstruction with adversarial training in an end-to-end neural work architecture design.

III. Problem Definition

In this section, let us formally define our migratable disaster damage assessment problem in social media sensing. We first define a few key terms used in the problem formulation.

Definition 1: **Source Disaster Event** (S): we define S to represent the source disaster event, where a set of high-quality training data (i.e., labeled imagery posts of damage severity) is available for the damage assessment task.

Definition 2: **Target Disaster Event** (T): we define T to represent the target disaster event, where the training data is not available.

Definition 3: **Disaster-related Social Media Images** (X): we define X to be the set of images posted on social media (e.g., Twitter) during a disaster event, where each image captures a specific scene of the disaster event. We define $X^S = \{X_1^S, X_2^S, ..., X_A^S\}$ and $X^T = \{X_1^T, X_2^T, ..., X_B^T\}$ to represent the set of images collected from the source and target event, respectively. A and B represent the number of social media images from source and target event, respectively.

Definition 4: **Event-specific Damage Characteristics**: we define event-specific damage characteristics to represent the visual characteristics (e.g., damage types, object layouts, and color distributions) of the social media images that are unique for a specific disaster event. For example, the four disaster events in Figure 1 clearly show the distinct event-specific damage characteristics in terms of dominant colors (e.g., grey of collapsed concrete in (A) vs. brown of broken wooden beam in (B)) and damage types (e.g., flood damage in (C) vs. wind damage in (D)).

Definition 5: **Latent Feature Vector** (V): we define V to be a latent feature vector, which represents a set of features (e.g., structural damage, landscape change) that indicate the the damage severity in a social media image. In addition, we define $V^S = \{V_1^S, V_2^S, ..., V_A^S\}$ and $V^T = \{V_1^T, V_2^T, ..., V_B^T\}$ to represent the latent feature vectors for the social media images from the source and target event, respectively.

Definition 6: **Damage Severity Level** (Y): we define the damage severity level Y to represent the severity of the damage captured in a disaster related social media image. For example, as proposed in [8], we can categorize the damage severity into three levels: severe damage, mild damage, no damage. In addition, we define $Y^S = \{Y_1^S, Y_2^S, ..., Y_A^S\}$ and $Y^T = \{Y_1^T, Y_2^T, ..., Y_B^T\}$ to represent the damage severity levels of images in the source and target event, respectively.

The goal of our migratable disaster damage assessment problem in social media sensing is to correctly identify the damage severity of affected areas reported in social media images from the target event that does not have a training dataset. Given the above definitions, we formally define our problem as follows:

$$\arg\max_{\widehat{Y_b^T}} \Pr(\widehat{Y_b^T} = Y_b^T \mid X^S, X^T, Y^S), \quad \forall 1 \leq b \leq B \tag{1}$$

where $\widehat{Y_b^T}$ is the *estimated* damage severity level for the b^{th} social media image collected from the the target event T. This problem is challenging due to the discrepancy on event-specific damage characteristics between the source and target events and the absence of training data for the target event. In this paper, we develop a SocialTrans scheme to address these challenges, which is elaborated in the next section.

IV. Solution

SocialTrans is a hybrid deep transfer learning framework to address the migratable disaster damage assessment problem defined in the previous section. In particular, it consists of two modules: 1) *Adversarial Co-Training Transfer Learning (ATLL)* and 2) *Multitask Joint Network Optimization (MJNO)*. The ATLL module designs an adversarial co-training deep transfer learning architecture that enables the effective damage assessment model migration between the source and target events. The *MJNO* module learns the optimal instances of all neural networks in the ATLL module to achieve the desirable damage assessment accuracy in the target disaster event.

A. Adversarial Co-Training Transfer Learning (ATLL)

The adversarial co-training transfer learning network design consists of four neural networks: a mapping network (*MN*), a discriminator network (*DN*), a reconstruction network (*RN*), and a classification network (*CN*). An overall architecture of the adversarial co-training transfer learning design is shown in Figure 2. In general, the *MN*, *DN*, *RN*, and *CN* work collaboratively to learn a migratable damage assessment model without using any training data from the target event. In particular, the *MN* is first used to map the images from both source and target events into latent feature vectors. The *DN* is then

used to examine the difference between the critical damage features extracted from source and target events. The goal is to regulate the *MN* so that it can effectively extract a critical set of damage features (e.g., structural damage, landscape change) shared between both source and target events. Meanwhile, the *RN* is used to translate the latent feature vectors generated by MN back to the images where the objective is to regulate the MN to ensure the critical damage features are successfully captured by the extracted latent feature vectors. Finally, the *CN* is used to classify the damage severity of input images using the latent feature vectors generated by *MN*. To the best of our knowledge, the *ATLL* is the first adversarial co-training architecture that integrates the deep image reconstruction (e.g., RN), the adversarial training (e.g., DN), and the damage severity classification (e.g., CN) in an end-to-end neural work architecture to identify the critical damage features without using any training data from the target event.

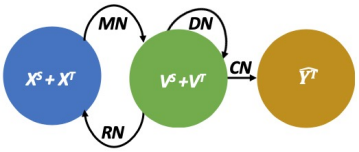


Figure 2. Overall of Adversarial Co-training Architectures

We formally define MN, DN, RN, and CN as follows:

Definition 7: **Mapping Network (*MN*)**: we define *MN* as a mapping network that maps social media images from both source and target events into latent feature vectors:

$$V^{S/T} = MN(X^{S/T}) \tag{2}$$

We show an example of the MN in (A) of Figure 3. It contains an ImageNet pre-trained deep convolutional neural network with multiple trainable convolutional layers for visual feature extraction. This is done to ensure the mapping network is able to 1) segment the complex visual features for an input image, and 2) identify the critical damage features from the segmented visual features.

Definition 8: **Discriminator Network (*DN*)**: we define *DN* as a discriminator network to examine whether a social media image comes from the source or target event using the latent feature vector generated by the MN:

$$DN : \begin{cases} \mathbf{1} : V \in S \\ \mathbf{0} : V \in T \end{cases} \tag{3}$$

where *DN* returns "1" if V is extracted from a social media image that comes from source area and "0" if it comes from target area. We show an example of the DN in (B) of Figure 3. It consists of a flatten layer and a few dense layers, which are used to determine whether a image belongs to the source or target event using the damage features extracted by the MN.

Definition 9: **Reconstruction Network (*RN*)**: we define *RN* as a reconstruction network that reconstructs social media images $X^{S/T}$ using the damage features in latent feature vectors $V^{S/T}$ as follows:

$$X^{S/T} = RN(V^{S/T}) \tag{4}$$

We show an example of the RN in (C) of Figure 3. It consists of a set of convolutional and upsampling layers, which are designed to convert the latent feature vectors generated by the MN back to the social media images. It regulates the MN to ensure that the critical set of visual features are successfully captured by the generated latent feature vectors.

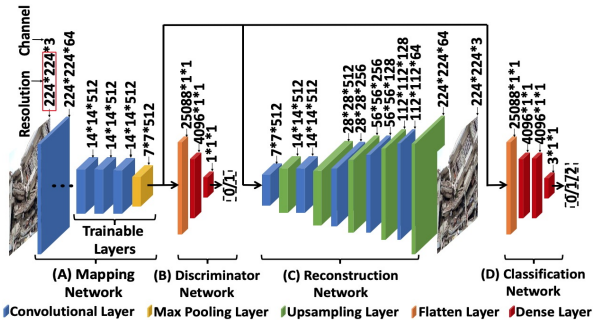


Figure 3. Illustrations of Network Architectures

Definition 10: **Classification Network (*CN*)**: we define *CN* as a classification network that estimates damage severity level of a social media image from the target event using the latent feature vector generated by the MN:

$$\widehat{Y^T} = CN(V^T) \tag{5}$$

We show an example of the CN in (D) of Figure 3. It consists of a flatten layer and multiple dense layers, which are used to estimate the damage severity level of an image using the damage features extracted by the MN.

B. Multitask Joint Network Optimization (MJNO)

Given the four network architectures above, our next question is how to learn the optimal instances of all networks to maximize the accuracy of the estimated damage severity level of images from the target event. To address this question, we define three sets of loss functions in our framework. In particular, we first consider the adversarial loss for the MN and DN as follows:

$$\begin{aligned} \mathcal{L}_{MN}^{Adv} &: |||\mathbf{0} - DN(MN(X^S))||_2 + ||\mathbf{1} - DN(MN(X^T))||_2 \\ \mathcal{L}_{DN}^{Adv} &: |||\mathbf{1} - DN(MN(X^S))||_2 + ||\mathbf{0} - DN(MN(X^T))||_2 \end{aligned} \tag{6}$$

where $\mathcal{L}_{MN}^{Adv}$ and $\mathcal{L}_{DN}^{Adv}$ represents the adversarial loss function for the MN and DN, respectively. X^S and X^T represent the social media images collected from source and target event, respectively. The goal of this loss function design is to set adversarial objectives for MN and DN so that DN can effectively regulate the MN to extract the critical damages features shared by both source and target events. On one hand, the $\mathcal{L}_{DN}^{Adv}$ is used to ensure DN can clearly identify whether an image comes from the source or target area using the extracted latent feature vector $X^{S/T}$. On the other hand, $\mathcal{L}_{MN}^{Adv}$ is used to ensure that MN can generate the critical damage features shared by both source and target areas so that DN is unable to distinguish them. Moreover, we further consider the

reconstruction loss for the MN and RN as follows:

$$\begin{aligned}&\mathcal{L}^{Rec}_{MN,RN}:\\&\mathcal{L}_{\text{pixel}}(X^S, RN(MN(X^S))) + \mathcal{L}_{\text{pixel}}(X^T, RN(MN(X^T)))\end{aligned} \tag{7}$$

where $\mathcal{L}^{Rec}_{MN,RN}$ represents the reconstruction loss function for MN and RN. $\mathcal{L}_{\text{pixel}}$ indicates the pixel-wise mean square error (MSE) loss [29] that measures the pixel-wise RGB value differences between the original and reconstructed images. The goal of this loss function is to check if the critical damage features shared by both source and target events are successfully captured by the MN. In particular, the latent feature vectors $MN(X^S)$ and $MN(X^T)$ generated by the MN are translated back to the social media images through RN and compared with the original ones to ensure the content is consistent (i.e., comparing $RN(MN(X^S))$ and $RN(MN(X^T))$ with X^S and X^T, respectively).

After defining the reconstruction loss function, our next question is how to accurately classify the damage severity of a social media image using the latent feature vector generated by the mapping network. To address this question, we define the classification loss $\mathcal{L}^{Cla}_{MN,CN}$ for the MN and CN as follows:

$$\mathcal{L}^{Cla}_{MN,CN} = \mathcal{L}_{\text{cross-entropy}}(Y^S, CN(MN(X^S))) \tag{8}$$

where Y^S is set of the ground-truth damage severity levels for the social media images X^S from the source event. $\mathcal{L}_{\text{cross-entropy}}$ indicates the cross entropy loss [30] that measures the difference between the real and estimated damage severity level of an image. The goal of this loss function is to check if CN can precisely classify the damage severity level of an image using the latent feature vectors captured by MN. Recall that the $\mathcal{L}^{Rec}_{MN,RN}$ in Equation (7) ensures that MN captures the critical damage features shared by both source and target events. The objective of this loss function is to ensure the CN can accurately assess damage severity of images from the target event by leveraging the identified features captured by MN. In particular, CN estimates the damage severity level $CN(MN(X^S))$ of an image using the latent feature vector $MN(X^S)$ and compares the estimation with the ground-truth label in Y^S of the source event.

We then combine the above three set of loss functions to derive the final loss $\mathcal{L}^{All}_{MN,RN,CN}$ for the generative networks (i.e., MN, RN, CN) and the final loss $\mathcal{L}^{All}_{MN,RN,CN}$ for the discriminator network (i.e., DN) to jointly optimize the objectives of our SocialTrans framework as follows:

$$\begin{aligned}&\mathcal{L}^{All}_{MN,RN,CN}: \mathcal{L}^{Adv}_{MN} + \mathcal{L}^{Rec}_{MN,RN} + \mathcal{L}^{Cla}_{MN,CN}\\&\mathcal{L}^{All}_{DN}: \mathcal{L}^{Adv}_{DN}\end{aligned} \tag{9}$$

Using the above loss functions, we can learn the optimal instances (i.e., MN^*, DN^*, RN^*, and CN^*) of all networks using the Adaptive Moment Estimation (ADAM) optimizer [31]. Finally, we use MN^* and CN^* to estimate the damage severity level for all input social media images X^T from the *target event* as follows:

$$\widehat{Y^T} = CN^*(MN^*(X^T)) \tag{10}$$

V. Evaluation

A. Dataset

In our evaluation, we use a real-world dataset collected from Twitter with ground-truth labels on damage severity [8]. In particular, the dataset consists of social media images collected from four different natural disaster events: Typhoon Ruby in Philippines (2014), Nepal Earthquake (2015), Ecuador Earthquake (2016), and Hurricane Matthew in USA (2016). These disaster events have a diversified set of damage characteristics (e.g., damage types, affected areas, and building structures), which creates a challenging evaluation scenario to study the migratable disaster damage assessment problem. In addition, the ground-truth damage severity level of each social media image is manually classified by human annotators into three categories (i.e., severe, mild, no). We keep the ratio of training to testing data as 3:1, the same as [8]. In our evaluation, we study a diversified set of source and target event combinations: 1) the source and target events belong to the same type of disaster (e.g., Ecuador and Nepal Earthquakes); 2) the source and target events belong to different but similar types of disasters (e.g., Typhoon Ruby and Hurricane Matthew); 3) the source and target events are completely different types (e.g., Ecuador Earthquake and Hurricane Matthew).

B. Baselines and Evaluation Metrics

We compare SocialTrans with a set of representative disaster damage assessment baselines.

- **InceptionNet [32]**: a widely used deep neural network that integrates convolution factorization to boost the learning process of the damage severity assessment.
- **DenseNet [33]**: a deep learning based disaster damage assessment framework that incorporates a feed-forward mechanism to achieve dense connections between convolutional layers.
- **MobileNet [34]**: a depthwise separable convolutional network that ensures a fast convergence of the learned damage severity classification model.
- **VGG16/19 [9]**: a representative deep learning approach that utilizes a stack of recursive convolutional operations to boost the classification accuracy. In our experiment, we consider two versions of VGG with different numbers of convolution layers (VGG16, VGG19).
- **GTA [16]**: a generate to adapt (GTA) domain adaptation approach that utilizes generative adversarial networks to learn the symbiotic relationship between the source and target domains for effective model migration.
- **DRCN [15]**: a state-of-the-art deep transfer learning approach that jointly leverages both the deep reconstruction and classification networks (DRCN) for effective migratable classification tasks.
- **DANN [12]**: a recent domain adversarial neural networks (DANN) framework that adds adversarial training mechanisms to the pre-trained VGG model for effective migratable disaster damage identification.

Table I
PERFORMANCE COMPARISONS (*Same* EVENT TYPE)

		Ecuador Earthquake→ Nepal Earthquake			Nepal Earthquake→ Ecuador Earthquake		
Category	Algorithm	F1-Score	$\mathcal{K}$-Score	MCC	F1-Score	$\mathcal{K}$-Score	MCC
Random	Random	0.3355	0.0013	0.0014	0.3596	0.0367	0.0406
Deep Convolutional Network	InceptionNet	0.6184	0.3491	0.3713	0.6360	0.3602	0.3764
	DenseNet	0.6206	0.3616	0.4077	0.6382	0.3869	0.4037
	MobileNet	0.6053	0.3171	0.3817	0.7325	0.5116	0.5223
	VGG16	0.6776	0.4533	0.4585	0.6996	0.4903	0.5040
	VGG19	0.6842	0.4567	0.4620	0.6864	0.4834	0.4994
Domain Adaptation	GTA	0.5197	0.2009	0.2067	0.5394	0.2054	0.2134
	DRCN	0.6535	0.3904	0.4010	0.6315	0.3722	0.3867
	DANN	0.7061	0.4778	0.4922	0.8157	0.6441	0.6486
Our Model	**SocialTrans**	**0.7434**	**0.5455**	**0.5579**	**0.8289**	**0.6788**	**0.6863**

To ensure a fairness comparison, we use the same inputs to all compared schemes, which include i) the social media images from both source and target disaster events, and ii) the ground-truth damage severity labels from the source event only. In addition, we also consider the *Random* baseline, which estimates the damage severity level of an image by randomly choosing a severity level from all possible candidates. In our experiment, we implement our model using TensorFlow[1] and Keras[2] libraries and train our model using the NVIDIA GTX 1080 Ti GPUs. In our experiment, all hyper-parameters are optimized using the Adam optimizer. In particular, we set the learning rate to be 10^{-5}. We also set the batch size to be 32 and the model is trained over 2000 epochs.

We adopt three representative metrics for multi-class classification problem to evaluate the performance of all compared schemes. In particular, we use *F1-Score*, *Cohen's kappa Score ($\mathcal{K}$-Score)*, and *Matthews Correlation Coefficient (MCC)*. The reason we select *$\mathcal{K}$-Score* and *MCC* is because our dataset is imbalanced (i.e., the damage severity class distribution of our dataset is: *None*: 42.6%, *Mild*: 13.9%, and *Severe*: 43.5%). Intuitively, a higher *F1-Score*, *MCC*, or *$\mathcal{K}$-Score* indicates a better classification performance.

C. Evaluation Results

In the first set of experiments, we evaluate the performance of all compared schemes by choosing the source and target event to be the same type of disaster. In particular, we consider two earthquake events (i.e., *Ecuador Earthquake* and *Nepal Earthquake*). We choose one earthquake event to be the source event and the other earthquake event to be the target event. For example, Ecuador Earthquake→Nepal Earthquake indicates that we choose the *Ecuador Earthquake* as the source event and the *Nepal Earthquake* as the target event. The evaluation results are shown in Table I. We observe that the SocialTrans scheme consistently outperforms all compared baselines. For example, the performance gains of SocialTrans compared to the best-performing baseline (i.e., DANN) in the Ecuador Earthquake→Nepal Earthquake scenario on F1-Score, $\mathcal{K}$-Score, and MCC are 3.73%, 6.77%, and 6.57%, respectively. Such performance gains mainly come from the fact that the adversarial co-training deep transfer learning network architecture design in SocialTras enables an effective model migration from the source to the target disaster event.

In the second set of experiments, we evaluate the performance of all compared schemes in more challenging scenarios where the source and target disaster events are of different types. In particular, we consider two evaluation settings: 1) source and target events are from two different but similar disaster types (i.e., *Typhoon Ruby* and *Hurricane Matthew*) and 2) source and target disaster events are of completely different types (i.e., *Ecuador Earthquake* and *Hurricane Matthew*, *Typhoon Ruby* and *Nepal Earthquake*). The evaluation results are shown in Table II and Table III. We observe that SocialTrans consistently outperforms all baselines across different evaluation settings. For example, the performance gains achieved by SocialTrans over the best-performing baseline (i.e., DRCN) in the Ecuador Earthquake→Hurricane Matthew scenario on F1-Score, $\mathcal{K}$-Score, and MCC are 5.88%, 8.65%, and 8.58%, respectively. The results demonstrate the robustness of SocialTrans in solving the migratable disaster damage assessment problem over different source and target event combinations (even across disaster types).

In the third set of experiment, we plot the ROC curves of SocialTrans and three top-performing baselines (i.e., the ones with the highest F1-Scores in the each of the tables shown above). The results are presented in Figure 4. Given the space limit, we only present the ROC curves for three source and target event pairs (e.g., Ecuador Earthquake→Nepal Earthquake in Table I). The results in other scenarios are similar. We observe that SocialTrans continues to outperform the best-performing baselines with the highest AUC value when the classification threshold changes.

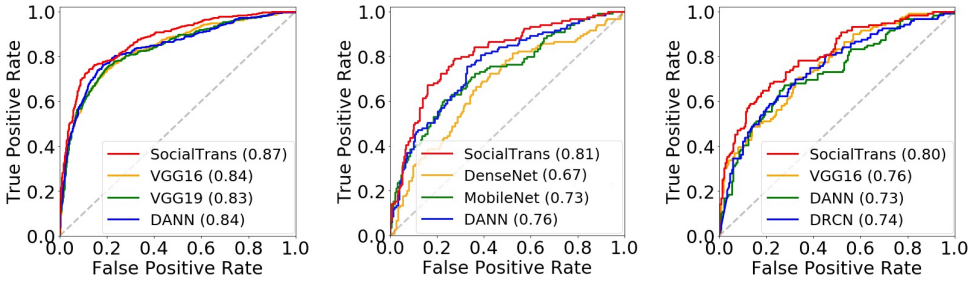
[1]https://www.tensorflow.org/
[2]https://keras.io/

Table II
PERFORMANCE COMPARISONS (*Similar* EVENT TYPE)

Category	Algorithm	Typhoon Ruby→ Hurricane Matthew F1-Score	$\mathcal{K}$-Score	MCC	Hurricane Matthew→ Typhoon Ruby F1-Score	$\mathcal{K}$-Score	MCC
Random	Random	0.3445	0.0544	0.0603	0.3493	0.0194	0.0206
Deep Convolutional Network	InceptionNet	0.4201	0.1759	0.2060	0.4397	0.0976	0.0995
	DenseNet	0.5210	0.2690	0.2883	0.3915	0.1458	0.1694
	MobileNet	0.5294	0.3130	0.3502	0.4337	0.1525	0.1635
	VGG16	0.4537	0.2042	0.2237	0.4156	0.1266	0.1356
	VGG19	0.5126	0.2653	0.2870	0.5542	0.2853	0.2894
Domain Adaptation	GTA	0.4453	0.1384	0.1504	0.4939	0.1305	0.1441
	DRCN	0.4873	0.1672	0.1710	0.4457	0.1013	0.1104
	DANN	0.5546	0.3385	0.3823	0.5662	0.2746	0.2822
Our Model	**SocialTrans**	**0.6722**	**0.4696**	**0.4847**	**0.6506**	**0.3865**	**0.3973**

Table III
PERFORMANCE COMPARISONS (*Different* EVENT TYPE)

Category	Algorithm	Ecuador Earthquake→ Hurricane Matthew F1-Score	$\mathcal{K}$-Score	MCC	Typhoon Ruby→ Nepal Earthquake F1-Score	$\mathcal{K}$-Score	MCC
Random	Random	0.3445	0.0151	0.0159	0.3373	0.0033	0.0034
Deep Convolutional Network	InceptionNet	0.4638	0.2114	0.2241	0.4638	0.2114	0.2241
	DenseNet	0.5180	0.2575	0.2606	0.5180	0.2575	0.2606
	MobileNet	0.4285	0.1826	0.2221	0.5060	0.2266	0.2302
	VGG16	0.5294	0.2559	0.2634	0.4939	0.2102	0.2116
	VGG19	0.5210	0.2188	0.2211	0.5421	0.2691	0.2695
Domain Adaptation	GTA	0.4705	0.1940	0.2286	0.4036	0.1115	0.1306
	DRCN	0.5798	0.2994	0.3302	0.4879	0.2215	0.2333
	DANN	0.5630	0.3003	0.3431	0.4457	0.1035	0.1448
Our Model	**SocialTrans**	**0.6386**	**0.3859**	**0.4160**	**0.5963**	**0.3499**	**0.3512**



(a) Ecuador Earthquake→Nepal Earthquake (*Same* Event Type) (b) Typhoon Ruby→Hurricane Matthew (*Similar* Event Type) (c) Ecuador Earthquake→Hurricane Matthew (*Different* Event Type)

Note that we compare SocialTrans with the three best-performing baselines with the highest F1-Score in the ROC curves of each scenario.

Figure 4. ROC Curves of SocialTrans and Best-Performing Baselines

VI. CONCLUSION

This paper presents a SocialTrans framework to solve the migratable disaster damage assessment problem in social media sensing. In particular, we develop a hybrid deep transfer learning framework to accurately identify the damage severity level of affected areas without using any training data from the target event. The results on real-world datasets from four different disaster events show that SocialTrans significantly outperforms state-of-the-art damage assessment baselines. We believe the unsupervised nature of SocialTrans makes it applicable to address similar data sparsity problems in

various social media sensing applications (e.g., misinformation tracking, disease outbreak detection) where the training data may not be available in the studied events of interest.

ACKNOWLEDGMENT

This research is supported in part by the National Science Foundation under Grant No. CNS-1845639, CNS-1831669, Army Research Office under Grant W911NF-17-1-0409. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, and L. Kaplan, "The age of social sensing," *Computer*, vol. 52, no. 1, pp. 36–45, 2019.

[2] D. Wang, T. Abdelzaher, and L. Kaplan, *Social sensing: building reliable systems on unreliable data*. Morgan Kaufmann, 2015.

[3] L. E. Charles-Smith, T. L. Reynolds, M. A. Cameron, M. Conway, E. H. Lau, J. M. Olsen, J. A. Pavlin, M. Shigematsu, L. C. Streichert, K. J. Suda *et al.*, "Using social media for actionable disease surveillance and outbreak management: a systematic literature review," *PloS one*, vol. 10, no. 10, 2015.

[4] A. I. J. T. Ribeiro, T. H. Silva, F. Duarte-Figueiredo, and A. A. Loureiro, "Studying traffic conditions by analyzing foursquare and instagram data," in *Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks*, 2014, pp. 17–24.

[5] L. Wu, Y. Zhi, Z. Sui, and Y. Liu, "Intra-urban human mobility and activity transition: Evidence from social media check-in data," *PloS one*, vol. 9, no. 5, p. e97010, 2014.

[6] D. Y. Zhang, Y. Huang, Y. Zhang, and D. Wang, "Crowd-assisted disaster scene assessment with human-ai interactive attention." in *AAAI*, 2020, pp. 2717–2724.

[7] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," *ACM Computing Surveys (CSUR)*, vol. 47, no. 4, pp. 1–38, 2015.

[8] D. T. Nguyen, F. Ofli, M. Imran, and P. Mitra, "Damage assessment from social media imagery data during disasters," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 569–576.

[9] X. Li, D. Caragea, H. Zhang, and M. Imran, "Localizing and quantifying damage in social media images," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018, pp. 194–201.

[10] F. Alam, M. Imran, and F. Ofli, "Image4act: Online social media image processing for disaster response," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 601–604.

[11] D. Zhang, Y. Zhang, Q. Li, T. Plummer, and D. Wang, "Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 1221–1232.

[12] X. Li, D. Caragea, C. Caragea, M. Imran, and F. Ofli, "Identifying disaster damage images using a domain adaptation approach."

[13] F. Alam, F. Ofli, M. Imran, and M. Aupetit, "A twitter tale of three hurricanes: Harvey, irma, and maria," *arXiv preprint arXiv:1805.05144*, 2018.

[14] D. Y. Zhang, C. Zheng, D. Wang, D. Thain, X. Mu, G. Madey, and C. Huang, "Towards scalable and dynamic social sensing using a distributed computing framework," in *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*. IEEE, 2017, pp. 966–976.

[15] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *European Conference on Computer Vision*. Springer, 2016, pp. 597–613.

[16] S. Sankaranarayanan, Y. Balaji, C. D. Castillo, and R. Chellappa, "Generate to adapt: Aligning domains using generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8503–8512.

[17] D. Wang, D. Zhang, Y. Zhang, M. T. Rashid, L. Shang, and N. Wei, "Social edge intelligence: Integrating human and artificial intelligence at the edge," in *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, 2019, pp. 194–201.

[18] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal, "On credibility estimation tradeoffs in assured social sensing," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1026–1037, 2013.

[19] D. Zhang, Y. Ma, C. Zheng, Y. Zhang, X. S. Hu, and D. Wang, "Cooperative-competitive task allocation in edge computing for delay-sensitive social sensing," in *2018 IEEE/ACM Symposium on Edge Computing (SEC)*. IEEE, 2018, pp. 243–259.

[20] Y. Zhang, D. Zhang, N. Vance, and D. Wang, "Optimizing online task allocation for multi-attribute social sensing," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2018, pp. 1–9.

[21] N. Vance, D. Zhang, Y. Zhang, and D. Wang, "Towards optimal incentive-driven verification in social sensing based smart city applications," in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2019, pp. 2700–2707.

[22] D. Zhang, Y. Ma, X. S. Hu, and D. Wang, "Towards privacy-aware task allocation in social sensing based edge computing systems," *arXiv preprint arXiv:2006.03178*, 2020.

[23] Y. Zhang, R. Zong, J. Han, H. Zheng, Q. Lou, D. Zhang, and D. Wang, "Transland: An adversarial transfer learning approach for migratable urban land usage classification using remote sensing," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 1567–1576.

[24] J. Choe, S. Park, K. Kim, J. Hyun Park, D. Kim, and H. Shim, "Face generation for low-shot learning using generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 1940–1948.

[25] Y. Zhang, R. Zong, J. Han, D. Zhang, T. Rashid, and D. Wang, "Transres: a deep transfer learning approach to migratable image super-resolution in remote urban sensing," in *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2020, pp. 1–9.

[26] G. Zhou, Z. Zeng, J. X. Huang, and T. He, "Transfer learning for cross-lingual sentiment classification with weakly shared deep neural networks," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 245–254.

[27] Y. Zhang, D. Zhang, and D. Wang, "On migratable traffic risk estimation in urban sensing: A social sensing based deep transfer network approach," *Ad Hoc Networks*, p. 102320, 2020.

[28] S. Min, M. Seo, and H. Hajishirzi, "Question answering through transfer learning from large fine-grained supervision data," *arXiv preprint arXiv:1702.02171*, 2017.

[29] M. S. Sajjadi, B. Scholkopf, and M. Hirsch, "Enhancenet: Single image super-resolution through automated texture synthesis," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4491–4500.

[30] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks." in *ICML*, vol. 2, no. 3, 2016, p. 7.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[32] C. Wang, D. Chen, L. Hao, X. Liu, Y. Zeng, J. Chen, and G. Zhang, "Pulmonary image classification based on inception-v3 transfer learning model," *IEEE Access*, vol. 7, pp. 146 533–146 541, 2019.

[33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks." in *CVPR*, vol. 1, no. 2, 2017, p. 3.

[34] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.